

8 de noviembre de 2007  
Los Polvorines, Provincia de Buenos Aires  
(Argentina)



*El ciclo del conocimiento en el entorno académico*

# Web Semántica y Bibliotecas Digitales

Dra. Eva M<sup>a</sup> Méndez Rodríguez  
[emendez@bib.uc3m.es](mailto:emendez@bib.uc3m.es)

Universidad Carlos III  
de Madrid



Dpto. Biblioteconomía y  
Documentación

## De qué vamos a hablar...

- ❏ “Problema” del conocimiento distribuido en la WWW: La Web hoy... Web 2.0, Web 3.0...
- ❏ Orígenes, dificultades y concepto de SW. ¿Qué es la Web Semántica?: una visión
- ❏ Implicaciones de la Web semántica para las BDs: Interoperabilidad, interdisciplinariedad, acceso global.
- ❏ Estándares y tecnologías de la Web semántica para las BDs.
- ❏ Conclusiones...



---

**“Problema” del conocimiento  
distribuido en la WWW: La Web Hoy...  
Web 2.0, Web 3.0...**

# La Web HOY

- ❏ Grandes cantidades de información de distinta naturaleza (recursos de alta calidad o profesionales con sitios informales, personales, de información poco fiable).
- ❏ PROBLEMA: organización, gestión y recuperación de esa información:
  - ✓ Edición/publicación de contenidos.
  - ✓ Accesos y modificación de la información.
  - ✓ Integración de recursos distribuidos en servicios de información o en servicios Web.
  - ✓ Valoración / Confianza de contenidos.
- ❏ Soluciones habituales → Insuficientes.



## La Web HOY: “efecto Alicia”

*¿Podrías decirme, por favor, qué camino he de tomar para salir de aquí?*

*Eso depende del lugar donde quieras ir -contestó el Gato.*

*Me da casi lo mismo el lugar -dijo Alicia.*

*En ese caso, no importa qué camino sigas -dijo el Gato.*

*Con tal de **llegar a alguna parte**... -añadió Alicia, tratando de explicarle.*

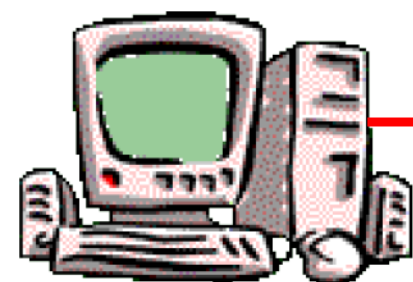
*¡Ah!... De eso no hay duda, **siempre que camines lo suficiente** -respondió el Gato”.*



Lewis Carroll

**Alicia en el país de las maravillas**

# La Web HOY: conocimiento distribuido



Bases de datos

Bibliotecas  
digitales

Buscadores

Colecciones de  
museos

E-Prints

Repositorios

Com Social

# La Web HOY: Tendencia *Amazoog*le



*El principio central que subyace a los gigantes nacidos en la Era de la Web 1.0, que han sobrevivido en la Era Web 2.0, parecen haber adoptado el poder de la Web para **explotar la inteligencia colectiva** (O'Reilly)*

Google™

YAHOO!

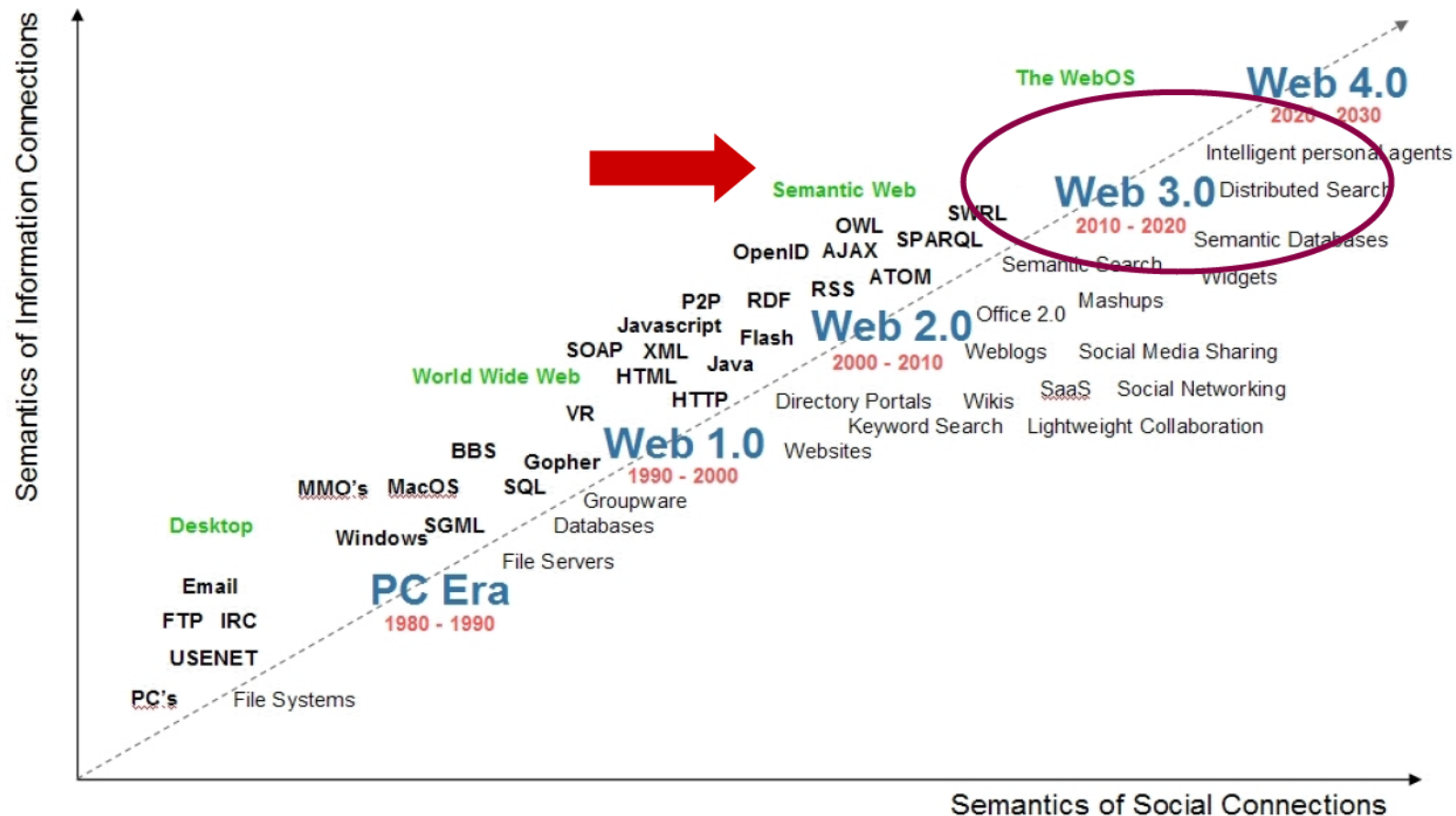


*“Glanceability” → permitir a los usuarios entender la información sin un esfuerzo cognitivo (Dempsey).*

ebay®

amazon.com™

# Evolución de la Web x.0: Web 3.0



Source: Radar Networks & Nova Spivack, 2007 – [www.radarnetworks.com](http://www.radarnetworks.com)

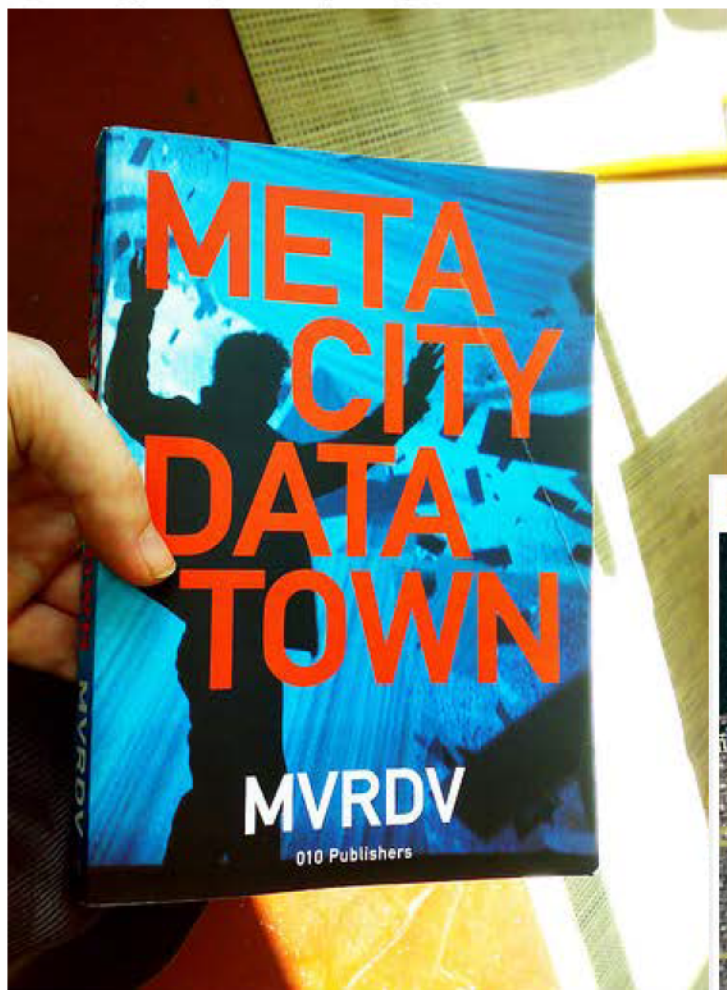
A través de IRSweb (Javier Martínez Méndez):

<http://irsweb.blogspot.com/2007/03/cronologa-de-la-web.html>



# SW y Web 2.0: dctagging y geotagging

ADD TO FAVES ADD NOTE BLOG THIS ALL SIZES ORDER PRINTS



by [Rev Dan Catt](#)

Rev Dan Catt's  
photostream



1,329  
photos

[View as  
slideshow](#)

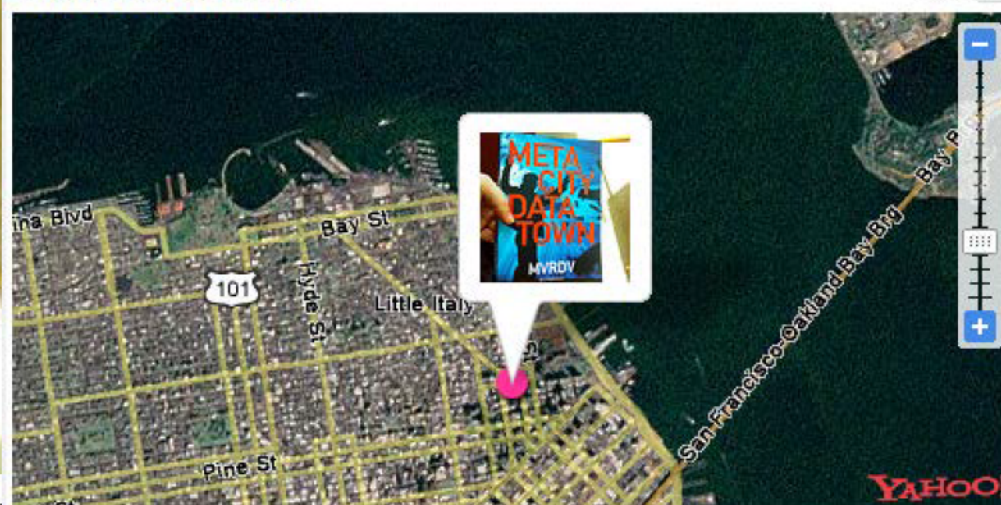
more browse more

Tags

- metacitydatatown
- book
- mrvdv

[View Rev Dan Catt's map](#)

Close



undefined

37°47' 41" N, 122°24' 06" W


[Eric](#) from [Stamen](#) directed me towards this book, out on the desk for normal place in the Stamen Library.

# La Web en el FUTURO (próximo)

- ❏ Continuará la mezcla de calidades informativas:
  - ✓ Se mantendrán sitios personales
  - ✓ Escaparates comerciales con un propósito real
- ❏ Los ordenadores utilizarán información estructurada de forma mucho más eficaz.
- ❏ Se necesita la integración de datos para permitir el procesamiento automático.
- ❏ Todavía mucha más participación/interacción del usuario.
- ❏ **La Web será “Semántica”...SW**
  - “...is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation”*


Tim Berners-Lee, James Hendler, Ora Lassila. The Semantic Web.  
*Scientific American* (Mayo 2001).

# Información distribuida: Retos

 Dotar a la Web de más **semántica** y de mayor flexibilidad para trabajar con información heterogénea y “dispersa”



**Web Semántica (SW)**

 **Acceso** a información distribuida (SOA)



**Servicios Web (WS)**



---

**Orígenes, dificultades y concepto de SW. ¿Qué es la Web Semántica? La Web semántica una Web +bibliotecaria.**



# Orígenes de la Web Semántica

- ❏ Idea original de la Web ([Tim B. Lee](#), 1989).
- ❏ Berners-Lee: *Metadata Architecture* (1997); *Weaving the Web* (1999).
- ❏ Definición formal (2001):



*La Web Semántica es una extensión de la Web actual en la que la información se presenta con un **significado** bien definido, permitiendo a los ordenadores y a las personas trabajar conjuntamente.*

## Dificultades para entender la SW

- ❏ Web semántica = Web + semántica
- ❏ Intangibilidad del concepto (Cognitivismo, Inteligencia Artificial, Documentación).
- ❏ El entusiasmo y la proyección de sus principales promotores (Tim Berners-Lee y James Hendler)
- ❏ Un conjunto interminable de siglas, acrónimos y conceptos:
  - ✓ RDF, XML, DCMI, SVG, DALM, OIL, OWL, SOAP, WSDL, etc.
  - ✓ Metadatos, ontologías, lógica, inferencia, etc.

# ¿Qué es la Web Semántica?

- ❏ La SW es “muchas cosas” para “muchacha gente”...
- ❏ Conjunto de tecnologías para la organización y representación del conocimiento digital que añaden semántica interpretable por las máquinas.
- ❏ Objetivos:
  - ✓ Proporcionar un acceso inteligente a la información heterogénea y distribuida en la WWW, posibilitando a los agentes de software mediar entre las necesidades de los usuarios y los recursos de información disponibles.
  - ✓ Creación de una **Web de datos** con significado (información/conocimiento), de tal forma que un programa de ordenador pueda aprender tanto acerca de lo que quieren decir los datos, como acerca de la información necesaria para procesarlos.
- ❏ Procesamiento global e interoperable de la información Web (marcado semántico XML + estructuras de metadatos).

# ¿Qué es la Web Semántica?

## Explicación gradual:

La Web es un conjunto de recursos y enlaces

Estos recursos y enlaces están identificados por URIs

Los recursos pueden tipificarse

Para un usuario, esto convierte a la web en algo más interesante

Las máquinas, por su parte tienen que ser capaces de procesar información legible por máquina, permitiendo trabajar personas y máquinas a la vez e intercambiar el conocimiento de una forma más eficaz.

***"The bane of my existence is doing things that I know the computer could do for me." (Dan Connolly.W3C)***

# SW: una visión desde la Documentación

- ❏ No es la W2: es la Web.
- ❏ Implica crear una interfaz común a todos los conceptos y bases de conocimiento terminológico en la Web.
- ❏ Sistema de información de acceso integrado de datos, términos y conceptos (=información).
- ❏ Aunar distintos recursos que existen en diferentes sistemas o servicios de información digital basados en Web (diccionarios, tesauros, ontologías, esquemas de clasificación, etc): NKOS (*Network Knowledge Organization Systems/Services*).
- ❏ Facilitar la búsqueda por materias.
- ❏ CBU (Control Bibliográfico Universal) → CWU (Control Web Universal) → Semantic Web (*global knowledge exchange*).





---

**Implicaciones de la Web semántica para las BDs:**  
**Interoperabilidad**  
**Interdisciplinariedad**  
**Acceso global**

## SW implica...

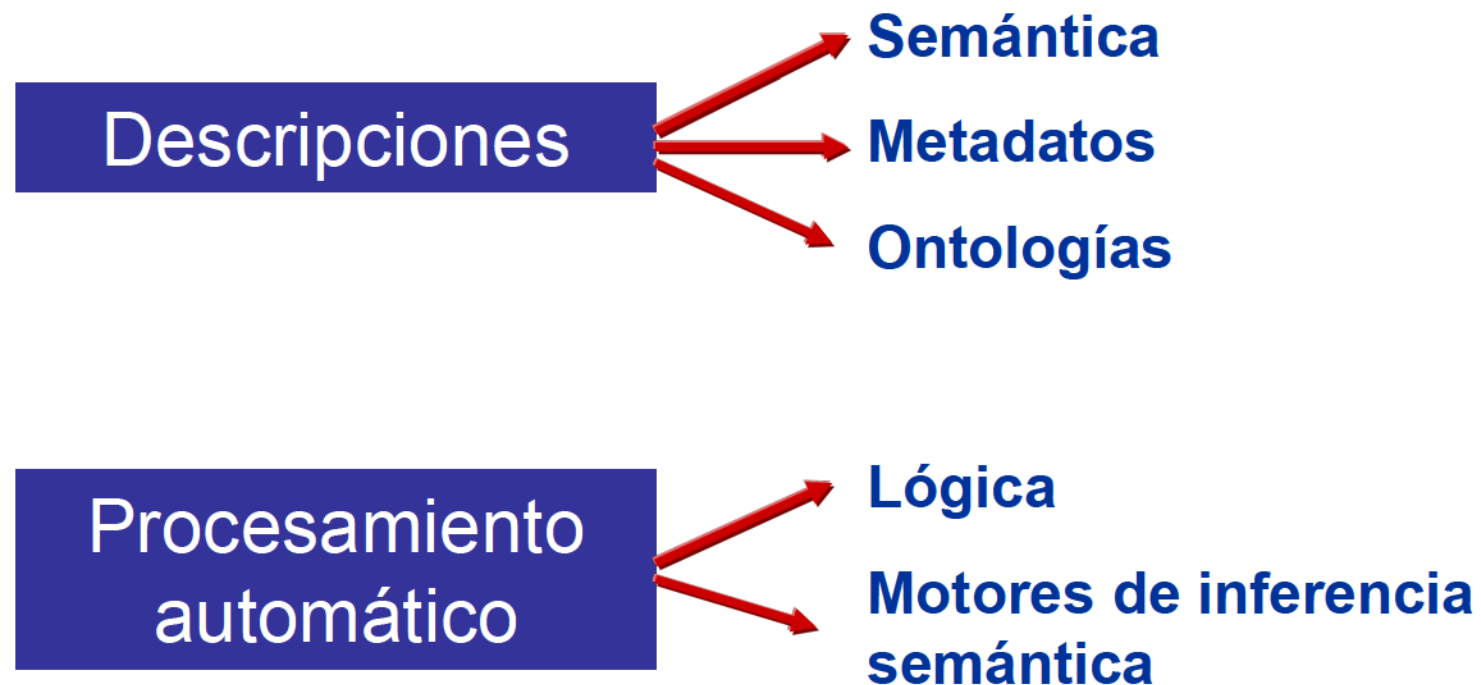
- ❏ Lenguajes formales como: DAML+OIL, RDF, OWL
- ❏ Metadatos y esquemas de metadatos y perfiles de aplicación codificados en RDF
- ❏ Ontologías: tanto específicas de un dominio, como ontologías de alto nivel de tal forma que se proporcionen relaciones entre ontologías para la recuperación, el intercambio y la integración de la información interdisciplinar en la Web.

## SW implica...

- ❏ Espacios de nombre (*xmlns*) para identificar un *esquemas* y cada elemento y cada concepto dentro de sus esquemas.
- ❏ Un amplio esquema conceptual que pueda mantener todas los tipos de información semántica y especificar las relaciones entre ellos (*ontologías*).
- ❏ Una definición de servicios que puedan interactuar con el sistema.



# SW implica...



# Semántica

- ❏ Significados legibles por máquina (codificar la “semántica de mundo”).
- ❏ XML (eXtended Markup Language):
  - ✓ Estándar de representación, metalenguaje para el intercambio de datos en la web.
  - ✓ Desde un punto de vista estrictamente informático: XML no añade semántica.
  - ✓ Por ello, necesita además: **metadatos y ontologías** (semántica).

# Metadatos

- ... **Datos.**
- ... Datos sobre datos.
- ... **Descripciones estructuradas** de un objeto de información web.
- ... Un **conjunto de elementos** y atributos para caracterizar la información.
- ... **Estándares** semánticos para describir objetos de datos discretos, que precisan de una semántica y estructura.

## Su finalidad es...

- ... Describir, recuperar y preservar la información Web.
- ... Fundamentar la recuperación de información eficaz en sistemas y servicios de información digital (a.k.a. bibliotecas digitales).

# Ontologías

- ❏ Gruber (1993) [Borst (1997)]: *Especificación sistemática [formal] de una conceptualización [compartida]*
- ❏ Una ontología es una herramienta conceptual que define un vocabulario común para compartir información dentro de un determinado dominio
- ❏ 80's: IA. para compartir y reutilizar conocimiento
- ❏ 90's: Ingeniería Web (ingeniería ontológica): descripciones semánticas explícitas de recursos
- ❏ 2000's: Web Semántica

# Ontologías

- ❏ Ontología (def. más difundida): Especificación de una conceptualización.
- ❏ Son conjuntos de...
  - ✓ ... términos,
  - ✓ ... **relaciones** entre los términos (conceptos),
  - ✓ ... que describen un dominio de aplicación concreto.
- ❏ Su objetivo es:
  - ✓ Creación de un diccionario de términos compartidos y comprensibles para diferentes aplicaciones y/o comunidades que permita la interoperabilidad.
  - ✓ Cualificar los términos para que representen conceptos.



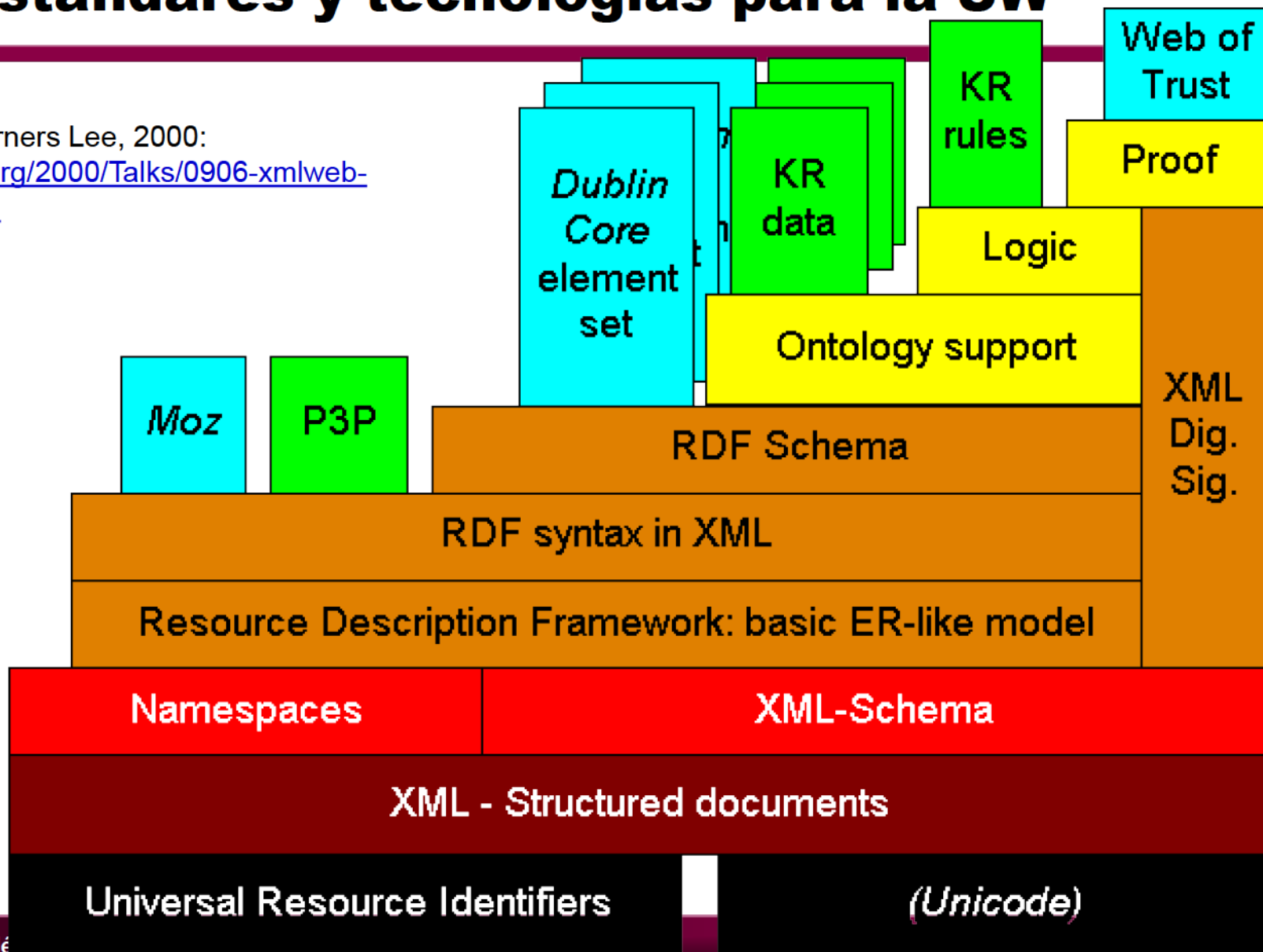
# **Estándares y tecnologías de la Web Semántica para las BDs**

**Entendiendo la sopa de letras... O poniendo más letras en la sopa**

# Estándares y tecnologías para la SW

Fuente: Tim Berners Lee, 2000:

<http://www.w3.org/2000/Talks/0906-xmlweb-tbl/slide9-0.html>





# Estándares y tecnologías para la SW

ISO 15836

DAML+OIL



OWL

FOAF

```
<epigraph>
  <ci>
    <part="N">
      "We are what suns and winds and waters make us."
    </part>
  </ci>
  <bibl rend="fv-sc">&#x2014;Landor.</bibl>
</epigraph>
```

**<?xml version="1.0">**




SVG





# Estándares y tecnologías para la SW

 **RDF (*Resource Description Framework*):**  
<http://www.w3.org/RDF>: 8 recomendaciones del  
W3C (10 feb 2004)


- ✓ [RDF/XML Syntax Specification \(Revised\)](#)
- ✓ [RDF Vocabulary Description Language 1.0: RDF Schema](#)
- ✓ [RDF Primer](#)
- ✓ [Resource Description Framework \(RDF\): Concepts and Abstract Syntax](#)
- ✓ [RDF Semantics](#)
- ✓ [RDF Test Cases](#)

# RDF: Características

- ❏ Basado en XML.
- ❏ Más que un mero formato de metadatos.
- ❏ Infraestructura de Descripción de Recursos.
- ❏ RDF propone un modelo de datos coherente y eficaz para representar (y/o crear lenguajes de representación) semántica.
- ❏ Influencia de los siguientes problemas:
  - ✓ PICS (*Platform for Internet Content Selection*)
  - ✓ Metadatos web (DC).
  - ✓ Sistemas de representación del conocimiento (tesauros, clasificaciones).
  - ✓ P3P (*Platform for Privacy Project*).

<http://www.w3.org/RDF>

# Estándares y tecnologías para la SW

 **OWL (Web Ontology Language):**  
<http://www.w3.org/2001/sw/WebOnt/>: 6 recomendaciones  
del W3C (10 febrero 2004)

- ✓ [OWL Web Ontology Language Overview](#)
- ✓ [OWL Web Ontology Language Guide](#)
- ✓ [OWL Web Ontology Language Reference](#)
- ✓ [OWL Web Ontology Language Semantics and](#)
- ✓ [Abstract Syntax](#)
- ✓ [OWL Web Ontology Language Test Cases](#)
- ✓ [OWL Web Ontology Language Use Cases and Requirements](#)  
+ 2 Notas: Presentación de la sintaxis en XML y analizador de  
OWL en RDF/XML

# Estándares y tecnologías para la SW

- ❏ DCMI (*Dublin Core Metadata Initiative*)
- ❏ DALM+OIL (*DARPA Agent Mark-Up Language- Ontology Inference Layer*)
- ❏ SKOS-Core: Esquema RDF para la codificación de tesauros y otros sistemas de organización del conocimiento.
- ❏ FOAF (*Friend of a Friend*)
- ❏ RSS (*RDF Site Summary*)
  
- ❏ ... Múltiples esquemas... (visión disciplinar):  
<http://www.schemaweb.info/>
- ❏ Lista de Ontologías (W3C-en desarrollo):  
<http://www.w3.org/2004/ontaria>



---

## **SW y bibliotecas**



# La Web necesita bibliotecari@s



*Lo que Internet necesita es un/a bibliotecario/a anticuado/a.*

*Encontrar lo que queremos en la Web debería de ser igual de fácil que encontrar un libro en una biblioteca.*

*Así será, si Eugenia [pongan aquí su nombre] consigue lo que quiere. Está trabajando para crear un **estándar** para etiquetar y catalogar información online –que incluye todos los 2.7 billones de páginas web... algo así como un sistema de información Decimal Dewey virtual, si lo preferís. Así, podrás gastar **menos tiempo en buscar, y más tiempo en utilizar la información** que necesitas.*

*Shhhhh... Estás en Internet.*

# La SW necesita Bib Digitales

- ❏ Cada vez más las organizaciones se dan cuenta de que necesitan bibliotecas digitales...  
→ ...*aunque no las llamen así.*
- ❏ La representación común de los datos y una arquitectura común reduce los costes (técnicos y sociales) y es más eficaz.
- ❏ Las “organizaciones de la memoria digital” (bibliotecas, archivos y museos) deben ser una fundamento clave para apoyar el desarrollo de la Web Semántica (E. Miller)
- ❏ Demostrar la validez de los servicios de agregación y/o sindicación de contenidos (RSS, OAI).

# Las Bib. Digitales necesitan SW

- ❏ Convergencia inminente de bibliotecas, museos y archivos digitales como “organizaciones de la memoria” (EDL)
- ❏ Convergencia de tecnologías facilitadoras (RDF, ontologías, etc.) para apoyar la convergencia de bibliotecas digitales, museos digitales y archivos digitales
- ❏ Desarrollo y uso de estándares de definición terminológica Web (SKOS, Zthes, etc.) → “webización” de tesauros y sistema de organización y representación del conocimiento.
- ❏ Las “BD” (Sistemas y servicios de información digital) necesitan identificar las áreas técnicas y sociales donde se solapan la SW y las Bibliotecas digitales.





---

**Conclusiones... Reflexiones... ¿y qué más?**

# Conclusiones

- ❏ Explosión de “colecciones”: El uso de las tecnologías de la SW para hacer efectiva la disponibilidad de contenidos.
- ❏ De una forma sencilla...
  - ...RDF es a la SW, lo que HTML a la Web.
- ❏ La consistencia en el acceso a la información por conceptos y su representación en RDF/XML, OWL... mejorará el acceso a la información en dos sentidos:
  - ✓ Una recuperación más precisa y eficaz
  - ✓ Un acceso a la información interdisciplinar (*cross-domain*)
- ❏ Necesaria investigación/inversión en metadatos y en organización del conocimiento.
- ❏ Principales problemas/objetivos:
  - ✓ **ACCESIBILIDAD**
  - ✓ **INTEROPERABILIDAD**
- ❏ Los bibliotecarios tenemos un papel muy importante en la construcción de la SW.

# SW



## Experiencia en:

- ✓ La creación y aplicación de metadatos.
- ✓ Organización de la información.
- ✓ Creación de estándares terminológicos (Tesauros, ontologías, etc.)

## Experiencia en:

- ✓ Lenguajes formales
- ✓ Procesamiento automático de la información.
- ✓ Inteligencia Artificial.

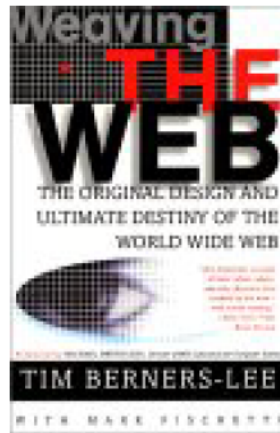
Compartir la experiencia adquirida en lenguajes de marcado, metadatos y vocabularios para la RI



**EL FUTURO DE LAS  
BIBLIOTECAS DIGITALES  
ESTÁ POR LLEGAR...**

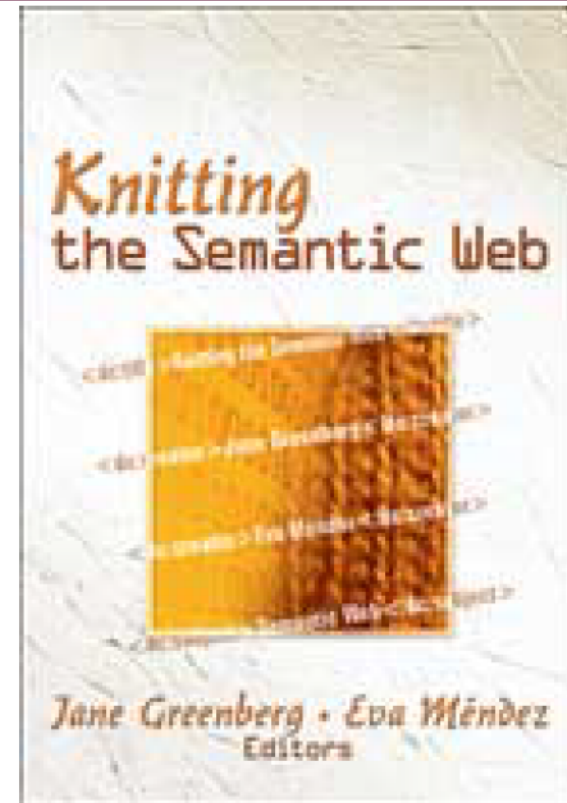


# Un libro... Web Semántica y Bibliotecarios...



Berners Lee (1999): *Weaving the Web* = Tejiendo la Web

- **Weaving**= como tejen las arañas (ants, buscadores)
- **Knitting**= como tejen las personas (los bibliotecarios)



Greenberg & Mendez (2007):  
*Knitting the Semantic Web* =  
Tejiendo la Web Semántica



# Gracias por su atención!!!

¿Discusión? ¿Preguntas?

[emendez@bib.uc3m.es](mailto:emendez@bib.uc3m.es)

